



UvA-DARE (Digital Academic Repository)

Good science, bad science: Questioning research practices in psychological research

Bakker, M.

[Link to publication](#)

Citation for published version (APA):

Bakker, M. (2014). Good science, bad science: Questioning research practices in psychological research

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

References

- Agresti, A., & Franklin, C. (2007). *Statistics: The art and science of learning from data*. Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*, 1-18. doi: 10.1037/A0021017
- American Psychological Association. (2001). *Publication Manual of the American Psychological Association (5th ed.)*. Washington, DC: American Psychological Association.
- American Psychological Association. (2010). *Publication Manual of the American Psychological Association (6th ed.)*. Washington, DC: American Psychological Association.
- American Psychological Association. (2012). Summary report of journal operations, 2011. *American Psychologist, 67*, 410-411. doi: 10.1037/a0028431
- Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory. *Current Directions in Psychological Science, 8*, 3-9. doi: 10.1111/1467-8721.00002
- Anisfeld, M. (1991). Neonatal imitation. *Developmental Review, 11*, 60-97. doi: 10.1016/0273-2297(91)90003-7
- Aron, A., Aron, E. N., & Coups, E. J. (2009). *Statistics for psychology (5th ed.)*. Upper Saddle River, NJ: Pearson.
- Asendorpf, J. B., Conner, M., Fruyt, F. D., Houwer, J. D., Denissen, J. J. A., Fiedler, K., . . . Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*, 108-119. doi: 10.1002/per.1919
- Augoustinos, M., Walker, I., & Donaghue, N. (2006). *Social cognition: An integrated introduction (2nd ed.)*. London, UK: Sage Publications Ltd.
- Bakker, M., Cramer, A. O. J., Matzke, D., Kievit, R. A., Van der Maas, H. L. J., Wagenmakers, E. J., & Borsboom, D. (2013). Dwelling on the past. *European Journal of Personality, 27*, 120-121. doi: 10.1002/per.1920

- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7, 543-554. doi: 10.1177/1745691612459060
- Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43, 666-678. doi: 10.3758/s13428-011-0089-5
- Bakker, M., & Wicherts, J. M. (in press). Outlier removal, sum scores, and the inflation of the Type I error rate in independent samples t tests. The power of alternatives and recommendations. *Psychological Methods*.
- Barber, T. X. (1976). *Pitfalls in human research. Ten pivotal points*. New York: Pergamon Press.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71, 230-244. doi: 10.1037//0022-3514.71.2.230
- Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. Chichester: John Wiley & Sons.
- Bates, D., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes (R package Version 0.999375-32) [Computer software]. Retrieved from <http://cran.r-project.org/web/packages/lme4/index.html>
- Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483, 531-533. doi: 10.1038/483531a
- Bem, D. J. (2000). Writing an empirical article. In R. J. Sternberg (Ed.), *Guide to publishing in psychology journals* (pp. 3-16). Cambridge, England: University Press.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407-425. doi: 10.1037/A0021524
- Benish, S. G., Quintana, S., & Wampold, B. E. (2011). Culturally adapted psychotherapy and the legitimacy of myth: A direct-comparison meta-analysis. *Journal of Counseling Psychology*, 58, 279-289. doi: 10.1037/A0023626
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, 57, 289-300.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer.
- Bergman, L. R. (1988). You can't classify all of the people all of the time. *Multivariate Behavior Research*, 23, 425-441.

- Berk, L. E. (2013). *Child development* (9th ed.). Boston: Pearson.
- Berle, D., & Starcevic, V. (2007). Inconsistencies between reported test statistics and p-values in two psychiatry journals. *International Journal of Methods in Psychiatric Research*, *16*, 202-207. doi: 10.1002/mpr.225
- Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2011). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology*, *97*, 613-636. doi: 10.1037/a0026739
- Blanton, H., Klick, J., Mitchell, G., Jaccard, J., Mellers, B., & Tetlock, P. E. (2009). Strong claims and weak evidence: reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, *94*, 567-582. doi: 10.1037/A0014665
- Bless, H., Fiedler, K., & Strack, F. (2004). *Social cognition: How individuals construct reality*. East Sussex, UK: Psychology Press.
- Board on Life Sciences National Research Council. (2003). *Sharing publication-related data and materials: Responsibilities of authors in the life sciences*. Washington, DC: National Academies Press.
- Bolstad, W. M. (2007). *Introduction to Bayesian statistics* (2nd ed.). Hoboken, NJ: Wiley.
- Bones, A. K. (2012). We knew the future all along: Scientific a priori hypothesizing is much more accurate than other forms of precognition. *Perspectives on Psychological Science*, *7*, 307-309. doi: 10.1177/1745691612441216
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons, Ltd.
- Brunner, E., & Munzel, U. (2000). The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal*, *42*, 17-25. doi: 10.1002/(Sici)1521-4036(200001)42:1<17::Aid-Bimj17>3.0.Co;2-U
- Buffardi, L. C., & Nichols, J. A. (1981). Citation impact, acceptance rate, and APA journals. *American Psychologist*, *36*, 1453-1456. doi: 10.1037/0003-066X.36.11.1453
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 1-12. doi: 10.1038/nrn3475
- Campbell, D. (1997). US Patent No. 75094728.
- Campbell, E. G., Clarridge, B. R., Gokhale, N. N., Birenbaum, L., Hilgartner, S., Holtzman, N. A., & Blumenthal, D. (2002). Data withholding in academic genetics - Evidence from

- a national survey. *Jama-Journal of the American Medical Association*, 287, 473-480. doi: 10.1001/jama.287.4.473
- Caperos, J. M., & Pardo, A. (2013). Consistency errors in p-values reported in Spanish psychology journals. *Psicothema*, 25, 408-414. doi: 10.7334/psicothema2012.207
- Card, N. A., Bosch, L., Casper, D. M., Wiggs, C. B., Hawkins, S. A., Schlomer, G. L., & Borden, L. M. (2011). A meta-analytic review of internalizing, externalizing, and academic adjustment among children of deployed military service members. *Journal of Family Psychology*, 25, 508-520. doi: 10.1037/A0024395
- Carey, B. (2011, November 2). Fraud case seen as a red flag for Psychology Research, *The New York Times*. Retrieved from <http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html>
- Ceci, S. J., & Walker, E. (1983). Private archives and public needs. *American Psychologist*, 38, 414-423.
- Cesario, J., Higgins, E. T., & Plaks, J. E. (2006). Automatic social behavior as motivated preparation to interact. *Journal of Personality and Social Psychology*, 90, 893-910. doi: 10.1037/0022-3514.90.6.893
- Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex*, 49, 609-610. doi: <http://dx.doi.org/10.1016/j.cortex.2012.12.016>
- Chambers, C. D., & Munafò, M. R. (2013, June 5). Trust in science would be improved by study pre-registration, *The Guardian*. Retrieved from <http://gu.com/p/3gbvn>
- Champely, S. (2009). *pwr: Basic functions for power analysis. R package, version 1.1.1.*
- Cliff, N. (1996). *Ordinal methods for behavioral data analysis*. Mahwah, NJ: Erlbaum.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, 65, 145-153. doi: 10.1037/h0045186
- Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolmann (Ed.), *Handbook of clinical psychology* (pp. 95-121). New York: McGraw-Hill.
- Cohen, J. (1990). Things I have learned (thus far). *American Psychologist*, 45, 1304-1312. doi: 10.1037/0003-066X.45.12.1304
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. doi: 10.1037/0033-2909.112.1.155
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003. doi: 10.1037/0003-066X.49.12.997

- Collet, D., & Lewis, T. (1976). The subjective nature of outlier rejection procedures. *Applied Statistics*, *25*, 228-237.
- Cooper, H., DeNeve, K., & Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods*, *2*, 447-452.
- Coursol, A., & Wagner, E. E. (1986). Effect of positive findings on submission and acceptance rates - A note on meta-analysis bias. *Professional Psychology-Research and Practice*, *17*, 136-137.
- Cromie, W. J. (1999). Mozart effect hits sour notes. Retrieved from <http://news.harvard.edu/gazette/1999/09.16/mozart.html>
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*: Routledge New York.
- Cumming, G. (2013). The new statistics: Why and how. *Psychological Science*. doi: 10.1177/0956797613504966
- Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., . . . Wilson, S. (2007). Statistical reform in psychology. *Psychological Science*, *18*, 230-232. doi: 10.1111/j.1467-9280.2007.01881.x
- Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology* (4th ed.). Harlow, UK: Pearson.
- De Groot, A. D. (1956). De betekenis van "significantie" bij verschillende typen onderzoek. *Nederlands Tijdschrift voor de Psychologie*, *11*, 398-409.
- Derogatis, L. R. (1994). Symptom Checklist 90-R *Administration, scoring, and procedures manual* (3rd ed.). Minneapolis, MN: National Computer Systems.
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*, 274-290. doi: 10.1177/1745691611406920
- Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *Plos One*, *7*. doi: 10.1371/journal.pone.0029081
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, *71*, 5-23. doi: 10.1037/0022-3514.71.1.5
- Efron, B. (1986). Why isnt everyone a Bayesian. *American Statistician*, *40*, 1-5. doi: 10.2307/2683111
- Eich, E. (2013). Business not as usual. *Psychological Science*. doi: 10.1177/0956797613512465

- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Emons, W. H. M., Sijtsma, K., & Meijer, R. R. (2007). On the consistency of individual classification using short scales. *Psychological Methods, 12*, 105-120. doi: 10.1037/1082-989x.12.1.105
- Epskamp, S., & Nuijten, M. B. (in preparation). *statcheck*: Extract statistics from articles and recompute p values. R package version 0.1.0.
- Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63*, 591-601. doi: 10.1037/0003-066X.63.7.591
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *Plos One, 4*. doi: 10.1371/Journal.Pone.0005738
- Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *Plos One, 5*, e10068. doi: 10.1371/journal.pone.0010068
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*, 891-904. doi: 10.1007/s11192-011-0494-7
- Farber, B. A., & Doolin, E. M. (2011). Positive regard. *Psychotherapy, 48*, 58-64. doi: 10.1037/A0022141
- Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17*, 120-128. doi: 10.1037/a0024445
- Fiedler, K. (2011). Voodoo correlations are everywhere. Not only in neuroscience *Perspectives on Psychological Science, 6*, 163-171. doi: 10.1177/1745691611400237
- Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α -error control to validity proper: Problems with a short-sighted false-positive debate. *Perspectives on Psychological Science, 7*, 661-669. doi: 10.1177/1745691612462587
- Field, A. (2013). *Discoverig statistics using IBM SPSS statistics* (4th ed.). London, UK: Sage Publication Ltd.
- Firebaugh, G. (2007). Replication data sets and favored-hypothesis bias. *Sociological Methods & Research, 36*, 200-209. doi: 10.1177/0049124107306663
- Fisher, R. A. (1958). *Statistical methods for research workers*. New York: Hafner.
- Francis, G. (2012a). The same old new look: Publication bias in a study of wishful seeing. *i-Perception, 3*, 176-178. doi: 10.1068/i0519ic

- Francis, G. (2012b). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, *19*, 151-156. doi: 10.3758/s13423-012-0227-9
- Francis, G. (2013). Publication bias in “Red, Rank, and Romance in Women Viewing Men” by Elliot et al. (2010). *Journal of Experimental Psychology: General*, *142*, 292-296. doi: 10.1037/a0027923
- Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics* (4th ed.). New York, N.Y.: Norton & Company.
- Freese, J. (2007). Replication standards quantitative social science - Why not sociology. *Sociological Methods & Research*, *36*, 153-172. doi: 10.1177/0049124107306659
- Friedlander, F. (1964). Type I and Type II bias. *American Psychologist*, *19*, 198-199. doi: 10.1037/h0038977
- Fugelsang, J. A., Stein, C. B., Green, A. E., & Dunbar, K. N. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory. *Canadian Journal of Experimental Psychology*, *58*, 86-95. doi: 10.1037/h0085799
- Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, *103*, 933-948. doi: 10.1037/A0029709
- Garcia-Berthou, E., & Alcaraz, C. (2004). Incongruence between test statistics and P values in medical papers. *BMC Medical Research Methodology*, *4*:13. doi: 10.1186/1471-2288-4-13
- Gibbons, J. D., & Chakraborti, S. (2003). *Nonparametric statistical inference* (Vol. 168). Boca Raton FL: CRC press.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*, 587-606. doi: 10.1016/j.socec.2004.09.033
- Green, D., & Rosenfeld, B. (2011). Evaluating the gold standard: A review and meta-analysis of the structured interview of reported symptoms. *Psychological Assessment*, *23*, 95-107. doi: 10.1037/A0021149
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*, 1-20. doi: 10.1037/h0076157

- Greenwald, A. G., Poehlman, T. A., Uhlman, E. L., & Banaij, M. R. (2009). Understanding and using the implicit association test III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17-41. doi: 10.1037/a0015575
- Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology, 68*, 155-165. doi: 10.1037/0022-006X.68.1.155
- Hall, J. A. (1998). How big are nonverbal sex differences? In D. J. Canary & K. Dindia (Eds.), *Sex differences and similarities in communication* (pp. 155-177). Mahwah, NJ: Erlbaum.
- Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on Anxiety and depression. *Psychological Bulletin, 137*, 940-958. doi: 10.1037/A0024355
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics*. New York: Wiley.
- Hayes, L. A., & Watson, J. S. (1981). Neonatal imitation - Fact or artifact. *Developmental Psychology, 17*, 655-660. doi: 10.1037//0012-1649.17.5.655
- Hewstone, M., Stroebe, W., & Jonas, K. (2012). *An introduction to social psychology*. West Sussex, UK: Wiley-Blackwell.
- Hoekstra, R., Finch, S., Kiers, H. A. L., & Johnson, A. (2006). Probability as certainty: Dichotomous thinking and the misuse of p values. *Psychonomic Bulletin & Review, 13*, 1033-1037. doi: 10.3758/BF03213921
- Hoekstra, R., Kiers, H. A. L., & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)? *Frontiers in psychology, 3*. doi: 10.3389/fpsyg.2012.00137
- Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. J. (in press). Robust misinterpretations of confidence intervals. *Psychonomic Bulletin & Review*.
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). London, UK: Pearson.
- Howitt, D., & Cramer, D. (2011). *Introduction to statistics in psychology* (5th ed.). London, UK: Pearson.
- Hubbard, R., & Ryan, P. A. (2000). The historical growth of statistical significance testing in psychology - and its future prospects. *Educational and Psychological Measurement, 60*, 661-681. doi: 10.1177/0013164400605001
- Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

- Hull, J. G., Slone, L. B., Meteyer, K. B., & Matthews, A. R. (2002). The nonconsciousness of self-consciousness. *Journal of Personality and Social Psychology*, *83*, 406-424. doi: 10.1037//0022-3514.83.2.406
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *Plos Medicine*, *2*, 696-701. doi: 10.1371/journal.pmed.0020124
- Ioannidis, J. P. A. (2008a). Effect of formal statistical significance on the credibility of observational associations. *American Journal of Epidemiology*, *168*, 374-383. doi: 10.1093/aje/kwn156
- Ioannidis, J. P. A. (2008b). Why most discovered true associations are inflated. *Epidemiology*, *19*, 640-648. doi: 10.1097/EDE.0b013e31818131e7
- Ioannidis, J. P. A. (2010). Meta-research: The art of getting it wrong. *Research Synthesis Methods*, *1*, 169-184.
- Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, *7*, 645-654.
- Ioannidis, J. P. A., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, *4*, 245-253. doi: 10.1177/1740774507079441
- Jaeger, R. G., & Halliday, T. R. (1998). On confirmatory versus exploratory research. *Herpetologica*, *54*, S64-S66.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524-532. doi: 10.1177/0956797611430953
- Jubb, M. (2008). *To share or not to share: Publication and quality assurance of research data outputs*. London, UK: Research Information Network.
- Judd, C. M., & Gawronski, B. (2011). Editorial comment. *Journal of Personality and Social Psychology*, *100*, 406. doi: 10.1037/0022789
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196-217. doi: 10.1207/s15327957pspr0203_4
- Keselman, H. J., Othman, A. R., Wilcox, R. R., & Fradette, K. (2004). The new and improved two-sample t test. *Psychological Science*, *15*, 47-51. doi: 10.1111/j.0963-7214.2004.01501008.x
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahnik, S., Bernstein, M. J., . . . Nosek, B. A. (in press). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*.

- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.
- Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes*, *56*, 28-55. doi: 10.1006/obhd.1993.1044
- Koepke, J. E., Hamm, M., Legerstee, M., & Russell, M. (1983). Neonatal imitation - 2 failures to replicate. *Infant Behavior & Development*, *6*, 97-102. doi: 10.1016/S0163-6383(83)80012-5
- Kraemer, H. C., Gardner, C., Brooks, J. O., & Yesavage, J. A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. *Psychological Methods*, *3*, 23-31.
- Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*, 299-312.
- Kyzas, P. A., Loizou, K. T., & Ioannidis, J. P. A. (2005). Selective reporting biases in cancer prognostic factor studies. *Journal of the National Cancer Institute*, *97*, 1043-1055. doi: 10.1093/Jnci/Dji184
- Langdrige, D., & Hagger-Johnson, G. (2009). *Introduction to research methods and data analysis in psychology* (2nd ed.). Harlow, UK: Pearson.
- LeBel, E. P., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. A., & Smith, C. T. (2013). PsychDisclosure.org: Grassroots support for reforming reporting standards in psychology. *Perspectives on Psychological Science*, *8*, 424-432. doi: 10.1177/1745691613491437
- Ledgerwood, A., & Sherman, J. W. (2012). Short, sweet, and problematic? The rise of the short report in psychological science. *Perspectives on Psychological Science*, *7*, 60-66. doi: 10.1177/1745691611427304
- Leggett, N. C., Thomas, N. A., Loetscher, T., & Nicholls, M. E. (2013). The life of p: 'Just significant' results are on the rise. *The Quarterly Journal of Experimental Psychology*, 1-16. doi: 10.1080/17470218.2013.863371
- Leman, P., Bremner, A., Parke, R. D., & Gauvain, M. (2012). *Developmental psychology*. London: McGraw Hill.
- Levelt Committee, Noort Committee, & Drenth Committee. (2012). Flawed science: The fraudulent research practices of social psychologist Diederik Stapel.

- Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against non-significant findings. *Communication Monographs*, *76*, 286-302. doi: 10.1080/03637750903074685
- Light, R. J., & Pillemer, D. B. (1984). *Summing up. The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*, 1181-1209. doi: 10.1037/0003-066X.48.12.1181
- Lorr, M., McNair, D. M., & Droppleman, L. F. (1992). *Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.
- Lucassen, N., Tharner, A., Van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Volling, B. L., Verhulst, F. C., . . . Tiemeier, H. (2011). The association between paternal sensitivity and infant-father attachment security: A meta-analysis of three decades of research. *Journal of Family Psychology*, *25*, 986-992. doi: 10.1037/A0025855
- Macdonald, P. (1999). Power, Type I, and Type III error rates of parametric and nonparametric statistical tests. *The Journal of Experimental Education*, *67*, 367-379. doi: 10.1080/00220979909598489
- Mahoney, M. J. (1976). *Scientist as subject: The psychological imperative*. Cambridge, MA, US: Ballinger Publishing Company.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, *1*, 161-175. doi: 10.1007/BF01173636
- Mahoney, M. J. (1979). Psychology of the scientist - Evaluative review. *Social Studies of Science*, *9*, 349-375. doi: 10.1177/030631277900900304
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*, 537-542. doi: 10.1177/1745691612460688
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, 50-60. doi: 10.1214/aoms/1177730491

- Marszalek, J. M., Barber, C., Kohlhart, J., & Holmes, C. B. (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills, 112*, 331-348. doi: 10.2466/03.11.pms.112.2.331-348
- Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *Quarterly Journal of Experimental Psychology, 65*, 2271-2279. doi: 10.1080/17470218.2012.711335
- Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods, 9*, 147-163. doi: 10.1037/1082-989X.9.2.147
- McCall, R. B., & Carriger, M. S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development, 64*, 57-79.
- McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2006). Lessons from the JMCB Archive. *Journal of Money Credit and Banking, 38*, 1093-1107. doi: 10.1353/mcb.2006.0061
- McCullough, B. D., & Vinod, H. D. (2003). Verifying the solution from a nonlinear solver: A case study. *American Economic Review, 93*, 873-892. doi: 10.1257/000282803322157133
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806-834.
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 75-78. doi: 10.1126/science.198.4312.75
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., . . . Read, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*, 128-165. doi: 10.1037/0003-066X.56.2.128
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166. doi: 10.1037/0033-2909.105.1.156
- Miller, J. (1991). Short report: Reaction time analysis with outlier exclusion: Bias varies with sample size. *The Quarterly Journal of Experimental Psychology Section A, 43*, 907-912. doi: 10.1080/14640749108400962
- Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*, 267-296. doi: 10.1037/A0021890

- Molenaar, D., Wicherts, J. M., & Borsboom, D. (2005). *Participant loss in psychological research and the problem of exclusion: A validity issue*. (Bachelor's thesis), University of Amsterdam, Amsterdam.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2009). *Introduction to the practice of statistics* (6th ed.). New York: W. H. Freedman and Company.
- Murayama, K., Pekrun, R., & Fiedler, K. (2013). Research practices that can prevent an inflation of false-positive rates. *Personality and Social Psychology Review*. doi: 10.1177/1088868313496330
- Murphy, J. R. (2004). Statistical errors in immunologic research. *Journal of Allergy and Clinical Immunology*, *114*, 1259-1264.
- Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment - Experimental-study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85-95. doi: 10.1080/00335557743000053
- Nature. (2006). A fair share. Editorial. *Nature*, *444*, 653-654.
- Nature. (2011). Devil in the details. To ensure their results are reproducible, analysts should show their workings. *Nature*, *470*, 305-306. doi: 10.1038/470305b
- Nelson, N., Rosenthal, R., & Rosnow, R. L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. *American Psychologist*, *41*, 1299-1301.
- Neuhauser, M., Losch, C., & Jockel, K. H. (2007). The Chen-Luo test in case of heteroscedasticity. *Computational Statistics & Data Analysis*, *51*, 5055-5060. doi: 10.1016/j.csda.2006.04.025
- Neuliep, J. W., & Crandall, R. (1990). Editorial bias against replication research. *Journal of Social Behavior and Personality*, *5*, 85-90.
- Neuliep, J. W., & Crandall, R. (1993). Reviewer bias against replication research. *Journal of Social Behavior and Personality*, *8*, 21-29.
- Newman, J., Rosenbach, J. H., Burns, K. L., Latimer, B. C., Matocha, H. R., & Rosenthal Vogt, E. (1995). An experimental test of "the Mozart effect": Does listening to his music improve spatial ability? *Perceptual and Motor Skills*, *81*, 1379-1387. doi: 10.2466/pms.1995.81.3f.1379
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*, 241-301. doi: 10.1037//1082-989x.5.2.241

- NIH. (2013). Final NIH statement on sharing research data. NOT-OD-03-032.
- Nolan, S. A., & Heinzen, T. E. (2007). *Statistics for the behavioral sciences*. New York, NY: Worth Publishers.
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science, 7*, 657-660. doi: 10.1177/1745691612462588
- Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation, 9*, 1-12.
- Pashler, H., Harris, C., & Coburn, N. (2011). *Elderly-related words prime slow walking*. Retrieved from <http://www.PsychFileDrawer.org/replication.php?attempt=MTU%3D>
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*, 531-536. doi: 10.1177/1745691612463401
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*, 528-530. doi: 10.1177/1745691612465253
- Petrocelli, J. V., Clarkson, J. J., Whitmire, M. B., & Moon, P. E. (2013). When ab not equal c - c ': Published errors in the reports of single-mediator models. *Behavior Research Methods, 45*, 595-601. doi: 10.3758/s13428-012-0262-5
- Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effect-Shmozart effect: A meta-analysis. *Intelligence, 38*, 314-323. doi: 10.1016/j.intell.2010.03.001
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data Is associated with increased citation rate. *Plos One, 2*. doi: 10.1371/journal.pone.0000308
- PLoS ONE. (2011). PLoS Editorial and Publishing Policies. 7. Sharing of Materials, Methods, and Data. Retrieved August 1, 2011, from <http://www.plosone.org/static/policies.action#sharing>
- PLoS ONE. (2014). Guidelines for reviewers. Retrieved January 9th, 2014, from <http://www.plosone.org/static/reviewerGuidelines>
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery, 10*, 712-781. doi: 10.1038/Nrd3439-C1

- Psychonomic Society. (2012). Psychonomic society guidelines on statistical issues. Retrieved Januari 9th, 2014, from http://www.psychonomic.org/pdfs/2012_Psychonomic_Society_Guidelines_on_Statistical_Issues.pdf
- Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task-performance. *Nature*, *365*, 611-611. doi: 10.1038/365611a0
- Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Pearson Assessment.
- Reason, J. (1990). *Human error*. Cambridge: Cambridge University Press.
- Reidpath, D. D., & Allotey, P. A. (2001). Data sharing in medical research: An empirical investigation. *Bioethics*, *15*, 125-134. doi: 10.1111/1467-8519.00220
- Richard, F., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*, 331-363. doi: 10.1037/1089-2680.7.4.331
- Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's 'Retroactive facilitation of recall' effect. *Plos One*, *7*, e33423. doi: 10.1371/journal.pone.0033423
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
- Rosenthal, R. (1978). How often are our numbers wrong? *American Psychologist*, *33*, 1005-1008. doi: 10.1037/0003-066X.33.11.1005
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638-641. doi: 10.1037/0033-2909.86.3.638
- Rosenthal, R., & Gaito, J. (1963). The interpretation of levels of significance by psychological researchers. *Journal of Psychology*, *55*, 33-38. doi: 10.1080/00223980.1963.9916596
- Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276-1284. doi: 10.1037/0003-066X.44.10.1276
- Rossi, J. S. (1987). How often are our statistics wrong? A statistics class exercise. *Teaching of Psychology*, *14*, 98-101. doi: 10.1207/s15328023top1402_8
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*: John Wiley and Sons Ltd.

- Rotton, J., Levitt, M. J., & Foos, P. (1993). Citation impact, rejection rates, and journal value. *American Psychologist*, *48*, 911-912. doi: 10.1037/0003-066X.48.8.911
- Rouder, J. N., Speckman, P. L., Sun, D., & Morey, R. D. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225-237. doi: 10.3758/PBR.16.2.225
- Rozeboom, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin*, *57*, 416-428.
- Ruscio, J., & Roche, B. (2012). Variance heterogeneity in published psychological research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *8*, 1-11.
- Samejima, F. (1997). The graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 85-100). New York: Springer.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110-114.
- Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS Journals. *Plos One*, *4*. doi: 10.1371/Journal.Pone.0007078
- Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the t test to departures from population normality. *Psychological Bulletin*, *111*, 352-360. doi: 10.1037/0033-2909.111.2.352
- Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, *17*, 551-566. doi: 10.1037/A0029487
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication Is neglected in the social sciences. *Review of General Psychology*, *13*, 90-100. doi: 10.1037/A0015108
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies. *Psychological Bulletin*, *105*, 309-316. doi: 10.1037//0033-2909.105.2.309
- Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *American Statistician*, *55*, 62-71. doi: 10.1198/000313001300339950
- Shadish, W. R., Doherty, M., & Montgomery, L. M. (1989). How many studies are in the file drawer - An estimate from the family marital psychotherapy literature. *Clinical Psychology Review*, *9*, 589-603.

- Shaffer, D. R., & Kipp, K. (2009). *Developmental psychology: Childhood and adolescence* (8th ed.). Belmont CA: Wadsworth.
- Shiffler, R. E. (1988). Maximum Z scores and outliers. *The American Statistician*, *42*, 79-80.
- Siegler, R. S., DeLoache, J. S., & Eisenberg, N. (2011). *How children develop* (3th ed.). New York: Worth publishers.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi: 10.1177/0956797611417632
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology*, *26*, 4-7.
- Simonsohn, U., Nelson, L., & Simmons, J. (in press). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General*.
- Spellman, B. (2012). Data, data, everywhere . . . especially in my file drawer. *Perspectives on Psychological Science*, *7*, 58-59. doi: 10.1177/1745691611432124
- Staudte, R. G., & Sheater, S. J. (1990). *Robust estimation and testing*. New York: Wiley.
- Sterba, S. K. (2006). Misconduct in the analysis and reporting of data: Bridging methodological and ethical agendas for change. *Ethics & Behavior*, *16*, 305-318.
- Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance - Or vice versa. *Journal of the American Statistical Association*, *54*, 30-34. doi: 10.1080/01621459.1959.10501497
- Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, *49*, 108-112. doi: 10.1080/00031305.1995.10476125
- Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99-110). New York: Wiley.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Stevens, J. (2001). *Applied multivariate statistics for the social sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Strasak, A. M., Zaman, Q., Marinell, G., Pfeiffer, K. P., & Ulmer, H. (2007). The use of statistics in medical research: A comparison of The New England Journal of Medicine and Nature Medicine. *American Statistician*, *61*, 47-55.
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, *7*, 670-688. doi: 10.1177/1745691612460687
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Needham Heights, MA Allyn & Bacon.
- Taleb, N. N. (2007). *The Black Swan*. New York: Random House.
- Tett, R. P., Meyer, J. P., & Roese, N. J. (1994). Applications of meta-analysis: 1987-1992. *International Review of Industrial and Organizational Psychology*, *9*, 71-112.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105-110. doi: 10.1037/h0031322
- Vacha-Haase, T., Nilsson, J. E., Reetz, D. R., Lance, T. S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology*, *10*, 413-425. doi: 10.1177/0959354300103006
- Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology Section A*, *47*, 631-650. doi: 10.1080/14640749408401131
- Vines, T. H., Albert, A. Y., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., . . . Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, *24*, 94-97. doi: 10.1016/j.cub.2013.11.014
- Voorbrood, C. (2010). *Archivering, beschikbaarstelling en hergebruik van onderzoeksdata in de psychologie*. The Hague, NL: Data Archiving and Networked Services.
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*, 274-290. doi: 10.1111/j.1745-6924.2009.01125.x
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779-804. doi: 10.3758/BF03194105
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment

- on Bem (2011). *Journal of Personality and Social Psychology*, *100*, 426-432. doi: 10.1037/a0022790
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*, 632-638. doi: 10.1177/1745691612463078
- Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, *29*, 350-362.
- Welch, B. L. (1947). The generalization of "student's" problem when several different population variances are involved. *Biometrika*, *34*, 28-35.
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*, 291-298. doi: 10.1177/1745691611406923
- Wicherts, J. M. (2008). Expectancy bias, statistical blunders, and the importance of reanalysis. Invited talk at the 23 IOPS summer conference, June 19, Enschede, The Netherlands.
- Wicherts, J. M., & Bakker, M. (2009). Sharing: guidelines go one step forwards, two steps back. *Nature*, *461*, 1053-1053. doi: 10.1038/4611053c
- Wicherts, J. M., & Bakker, M. (2012). Publish (your data) or (let the data) perish! Why not publish your data too? *Intelligence*, *40*, 73-76. doi: 10.1016/j.intell.2012.01.004
- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data Is related to the strength of the evidence and the quality of reporting of statistical results. *Plos One*, *6*. doi: 10.1371/journal.pone.0026828
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*, 726-728. doi: 10.1037/0003-066x.61.7.726
- Wicherts, J. M., & Johnson, W. (2009). Group differences in the heritability of items and test scores. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 2675-2683.
- Wicherts, J. M., Kievit, R. A., Bakker, M., & Borsboom, D. (2012). Letting the daylight in: Reviewing the reviewers and other ways to maximize transparency in science. *Frontiers in Computational Neuroscience*, *6*. doi: 10.3389/Fncom.2012.00020
- Wicherts, J. M., & Vorst, H. C. M. (2004). Modelpassing van de Verkorte Profile of Mood States en meetinvariantie over mannen en vrouwen. *Nederlands Tijdschrift voor de Psychologie*, *59*, 12-21. doi: 10.1007/BF03062320

- Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego: Academic Press.
- Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods. *American Psychologist*, *53*, 300-314. doi: 10.1037/0003-066X.53.3.300
- Wilcox, R. R. (2003). *Applying contemporary statistical techniques*. San Diego, CA: Academic Press.
- Wilcox, R. R. (2012). *Modern statistics for the social and behavioral sciences: A practical introduction*. Boca Raton, FL: CRC Press.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*, 80-83. doi: 10.2307/3001968
- Wilkinson, L. (1999). Statistical methods in psychology journals - Guidelines and explanations. *American Psychologist*, *54*, 594-604.
- Wilson, S., & MacLean, R. (2011). *Research methods and data analysis for psychology*. Berkshire, UK: McGraw-Hill.
- Wolins, L. (1962). Responsibility for raw data. *American Psychologist*, *17*, 657-658. doi: 10.1037/h0038819
- Woodin, E. M. (2011). A two-dimensional approach to relationship conflict: meta-analytic findings. *Journal of Family Psychology*, *25*, 325-335. doi: 10.1037/A0023791
- Woodley, M. A. (2011). The cognitive differentiation-integration effort hypothesis: A synthesis between the fitness indicator and life history models of human intelligence. *Review of General Psychology*, *15*, 228-245. doi: 10.1037/A0024348
- Yong, E. (2012). Replication studies: Bad copy. *Nature*, *485*, 298. doi: 10.1038/485298a
- Yuen, K. K. (1974). The two-sampled trimmed t for unequal population variances. *Biometrika*, *61*, 165-170. doi: 10.1093/biomet/61.1.165
- Zijlstra, W. P., van der Ark, L. A., & Sijtsma, K. (2007). Outlier detection in test and questionnaire data. *Multivariate Behavioral Research*, *42*, 531-555.
- Zimmerman, D. W. (1998). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. *The Journal of Experimental Education*, *67*, 55-68. doi: 10.1080/00220979809598344
- Zimmerman, D. W. (2004). A note on the influence of outliers on parametric and nonparametric tests. *Journal of General Psychology*, *121*, 391-401.

-
- Zimmerman, D. W., & Zumbo, B. D. (1992). Parametric alternatives to the student t test under violation of normality and homogeneity of variance. *Perceptual and Motor Skills, 74*, 835-844. doi: 10.2466/pms.1992.74.3.835